RESEARCH ARTICLE

Genetic diversity and DNA profiling of robusta germplasm using SSR markers

Shahril Ab Razak[1, *], Alny Marlynni Abd Majid[1], Noor Syahira Nasarudin[2], Anis Syazwani Kamarudin[1], Norsyahirah Azmi[1], Rupini Yesudasan[2], Umi Kalsum Bahari[2], Mohd Rani Awang[3]

[1]Biotechnology & Nanotechnology Research Centre, MARDI Headquarters, Selangor, Malaysia. [2]Industrial Crop Research Centre, MARDI Kluang, Johor, Malaysia. [3]Industrial Crop Research Center, MARDI Jerangau, Terengganu, Malaysia.

Received: September 18, 2024; **accepted:** January 9, 2025.

***Coffea canephora*, commonly known as Robusta coffee, accounts for nearly 40% of global coffee production and is valued for its resistance to diseases, pests, and adaptability to harsh environmental conditions. Despite its economic significance, the genetic diversity and population structure of Robusta coffee remain underexplored, posing challenges to breeding programs aimed at improving crop yield, quality, and resilience. This study aimed to assess genetic diversity and develop a DNA fingerprinting panel for Robusta germplasm using simple sequence repeat (SSR) markers. Twelve SSR markers were initially tested to evaluate 32 robusta accessions with 11 found to be polymorphic. The number of alleles per marker ranged from 2 to 12 with an average of 6.727. The polymorphic information content (PIC) values, which assess the discriminatory power of each marker, ranged from 0.065 (M809) to 0.841 (EST-SSR 029) with an average PIC of 0.5034. The genetic relationships among accessions were visualized using a UPGMA dendrogram, revealing three distinct genetic clusters. To enhance the efficiency of cultivar identification, an SSR fingerprinting panel was developed using five highly polymorphic SSR markers with PIC values above 0.6. This optimized panel demonstrated superior discrimination power with pairwise genetic distances ranging from 0.0 to 1.0, compared to 0.0 to 0.857 when using all 11 polymorphic markers. The SSR fingerprinting panel offers a cost-effective, reliable approach to the identification of Robusta accessions. This study underscores the utility of SSR markers in enhancing the genetic profiling and management of Robusta germplasm, supporting future breeding, conservation, and planting material purity evaluation efforts.**

Keywords: *Coffea canephora*; DNA fingerprinting; genetic variability; SSR marker.

[*]**Corresponding author:** Shahril Ab Razak, Biotechnology & Nanotechnology Research Centre, MARDI Headquarters, 43400 Serdang, Selangor, Malaysia. Email: shahrilf@mardi.gov.my.

Introduction

Coffea canephora, commonly known as Robusta coffee, is one of the most important species in the coffee industry, accounting for nearly 40% of global coffee production. Unlike *Coffea arabica*, which is more widely known, Robusta is preferred for its resistance to diseases, pests, and its adaptability to harsher environmental conditions. Its beans are used primarily in espresso and instant coffee blends due to their strong flavor and high caffeine content [1]. However, despite its economic significance, the genetic diversity and population structure of Robusta coffee remain underexplored compared to Arabica coffee. This gap in knowledge presents

challenges to the development of effective breeding programs that could improve crop yield, quality, and resilience. Understanding the genetic diversity of Robusta coffee is, therefore, critical for its conservation, sustainable use, and improvement.

Genetic diversity is the basis of adaptation and evolution, providing a buffer for populations against environmental pressures. A genetically diverse crop population is more likely to harbor traits that can withstand biotic and abiotic stressors, including diseases and changing climate conditions [2]. For crops like Robusta coffee, which faces increasing demand and environmental challenges, maintaining and enhancing genetic diversity is essential for long-term productivity. Although cultivar identification based on phenotypic traits remains important, relying solely on morphological traits can be inaccurate and inconsistent due to the influence of environmental and phenological factors. Additionally, this approach is often limited by the availability of a sufficient number of distinguishing traits. The best approach to assess the genetic diversity in plants is by using DNA marker, which allows researchers to assess variations at the molecular level. In this context, microsatellite markers or simple sequence repeats (SSRs) have emerged as a powerful tool for analyzing the genetic diversity of plants. These markers are highly polymorphic, co-dominant, and evenly distributed throughout the genome, making them ideal for DNA profiling [3]. Microsatellite markers have been widely used in plant genetics for over two decades due to their effectiveness in detecting genetic variation. They consist of short, tandemly repeated DNA sequences that vary in length among different individuals. The variation in these repeats can be used to differentiate between individuals, populations, or accessions, thereby providing insights into the genetic structure and diversity within a species [4]. One of the main advantages of using microsatellite markers in genetic studies is their ability to provide accurate and consistent approaches that can be used for DNA fingerprinting and germplasm characterization. DNA fingerprinting allows researchers to create unique genetic profiles for different accessions, enabling the identification of duplicate or mislabeled germplasm in gene banks. This is particularly important for maintaining the integrity of germplasm collections, which serve as valuable reservoirs of genetic diversity for breeding and conservation efforts [5].

This study aimed to assess the genetic diversity of Robusta germplasm using microsatellite markers. The findings would provide crucial information for future breeding programs, conservation efforts, and germplasm management initiatives. As global demand for coffee continues to rise and environmental challenges intensify, understanding and preserving the genetic diversity of Robusta coffee is essential to ensure the sustainability and resilience of this vital crop.

Materials and methods

Plant materials

Young leaves from 32 accessions of *Coffea canephora* were collected from the Malaysian Agricultural Research and Development Institute (MARDI) Coffea germplasm in Kluang, Johor, Malaysia that is in the southern region of the Malay Peninsula. The leaves were air-dried using silica gel before DNA extraction to ensure optimal preservation of the samples.

DNA extraction

DNA extraction was performed following the method described by Mace *et al*., with slight modifications in grinding and incubation time [6]. Approximately 1 g of leaf tissue was ground using a Tissue Lyser (Qiagen, Hilden, North Rhine-Westphalia, Germany) and incubated at 65°C for one hour in 600 µL of cetyltrimethylammonium bromide (CTAB) extraction buffer (2% CTAB, pH 8, 100 mM Tris-HCl, 20 mM EDTA, 1.4 M NaCl, 0.05% β-mercaptoethanol). The DNA was precipitated with an equal volume of cold isopropanol, then washed twice with 70% ethanol. The resulting DNA pellet was air-dried

Table 1. List of twelve SSR used in this study with their respective forward and reverse sequences and annealing temperature (Ta).

No	Marker ID	Forward	Reverse	Ta (°C)
1	M803	GGTACATGCTCCTTCCAAGA	TCTCTCTGTATCTCTTTACCTTCACC	58.3
2	M809	AGCAAGTGGAGCAGAAGAAG	CGGTGAATAAGTCGCAGTC	58.3
3	M811	TGGAGAAGGCTGTTGAAACC	GGCGTGAAGCAAAAAGGTAT	55.0
4	M826	CCGCACTCACACACTACTTCT	TCTTATCCTCTTCCATTGCTTC	58.3
5	SSRCa 018	GTCTCGTTTCACGCTCTCTC	ATTTTTGGCACGGTATGTTC	55.0
6	SSRCa 019	GGGTTAGATAGAGCAAGAATGA	CTGTGAAGGTGTGGAGTTTT	51.2
7	SSRCa 026	GAATCTGGTGGGCTTTGA	AAGGAGAGGGGAAGAAAATG	55.0
8	SSRCa 068	ATGTTGTTGGAGGCATTTTC	AGGAGCAGTTGTTGTTTTCC	58.3
9	SSRCa 083	TCCAACAACATTAAGCGTATTC	GACAAACCTGAGGGAAAAGA	58.3
10	SSRCa 087	TCACTCTCGCAGACACACTAC	GCAGAGATGATCACAAGTCC	55.0
11	EST-SSR 029	TTAACCTCCTGCCACACA	GCCCAAATAAATCCCTCCA	58.3
12	EST-SSR 054	GTTAGCCGTTGGTGATGGAA	TTGGTCGAGGGAGGAAGAAC	55.0

and re-suspended in 50 µL of TE-RNase buffer. DNA concentration was determined using the Thermo Labsystems Fluoroskan Ascent™ (Thermo Scientific, Waltham, Massachusetts, USA), and the integrity of the DNA was verified on a 0.8% agarose gel.

SSR selection and genotyping

Twelve SSR markers were selected from previous studies [7, 8]. Details of these selected markers with their primer sequences were shown in Table 1 with each forward primer being concatenated with an M13 sequence as outlined by Schuelke [9]. The polymerase chain reaction (PCR) was conducted in 10 µL reaction volume containing 1 µL of 10× Invitrogen PCR Buffer, 1 µL of 2.5 mM $MgCl_2$, 1 µL of genomic DNA, 1 µL of 2 µM dNTPs, 0.2 µL of 10 µM of each primer, and 0.1 µL of 1 U of Taq polymerase (Invitrogen, Waltham, Massachusetts, USA). The PCR was carried out using an Applied Biosystem GeneAmp (Thermo Fischer Scientific, Waltham, Massachusetts, USA) with the following profile of initial denaturation at 95°C for 2 minutes followed by 35 cycles of denaturation at 94°C for 30 seconds, annealing at 45–64°C for 30 seconds, and elongation at 70°C for 45 seconds. A final elongation step was performed at 70°C for 5 minutes. The resulting PCR products were analyzed using an ABI3730xl DNA Analyzer capillary array (Thermo Fischer Scientific, Waltham, Massachusetts, USA) with GeneScan™ 500 LIZ (Applied Biosystems, Waltham, Massachusetts, USA) as the DNA ladder.

Scoring and data analysis

Allele peaks in the electropherograms were scored and analyzed following the guidelines provided by Arif *et al*. [10]. The scored allele data were then transferred to Microsoft Excel and formatted using appropriate data input files. POWERMARKER (https://brcwebportal.cos.ncsu.edu/powermarker/) was employed to calculate the number of alleles, heterozygosity values, polymorphic information content (PIC), and genetic distance [11]. Subsequently, Nei's genetic distance was used to generate UPGMA dendrograms, which were constructed using MEGA7 (https://www.megasoftware.net/) [12].

Results and discussion

SSR marker characterization

The genetic diversity of Robusta germplasm was evaluated using 12 SSR markers. Out of twelve SSR markers, eleven were found to be polymorphic, while the remaining one SSR marker (EST-SSR 054) was monomorphic. The number of alleles per marker ranged from two (M809) to twelve (SSRCa 018) with an average of 6.727 alleles per marker. This variation in allele number highlighted the markers' capacity to capture genetic diversity within the Robusta germplasm with SSRCa 018 displaying the highest

Table 2. SSR marker characterization.

Marker ID	Major allele frequency	Allele No.	Gene diversity	Heterozygosity	PIC
SSRCa 018	0.337	12.000	0.827	0.851	0.811
SSRCa 068	0.871	3.000	0.228	0.257	0.208
SSRCa 019	0.416	9.000	0.712	0.726	0.669
SSRCa 087	0.753	6.000	0.420	0.172	0.404
M826	0.574	3.000	0.493	0.505	0.376
EST-SSR 029	0.238	9.000	0.857	0.871	0.841
SSRCa 026	0.268	9.000	0.842	0.863	0.823
SSRCa 083	0.426	11.000	0.771	0.426	0.751
M803	0.774	5.000	0.382	0.041	0.359
M809	0.965	2.000	0.067	0.070	0.065
M811	0.871	5.000	0.237	0.090	0.230
Mean	0.590	6.727	0.530	0.443	0.503

allelic richness, suggesting it was particularly useful for distinguishing between genotypes. The major allele frequency varied widely among the markers with values ranging from 0.238 (EST-SSR 029) to 0.965 (M809), reflecting the presence of both common and rare alleles in the germplasm. The average major allele frequency across all markers was 0.590, indicating a moderate level of allele frequency distribution. Heterozygosity, which indicated the proportion of individuals that were heterozygous at a given locus, ranged from 0.041 (M803) to 0.871 (EST-SSR 029) with an overall mean of 0.443. The PIC values, which assessed the discriminatory power of each marker, ranged from 0.065 (M809) to 0.841 (EST-SSR 029) with a mean PIC of 0.503. Markers such as EST-SSR 029, SSRCa 026, and SSRCa 018, which had PIC values above 0.8, were particularly informative, indicating their utility in genetic studies for Robusta coffee. Markers with high PIC values were valuable for distinguishing between genotypes and assessing genetic diversity, making them suitable for use in genetic mapping and association studies [13]. Markers such as M809 and M811 with low gene diversity, heterozygosity, and PIC values, contributed less to the overall genetic diversity assessment. These markers might be less useful for distinguishing between closely related genotypes but could still provide valuable information in specific breeding or conservation contexts. Meanwhile, the selection of highly informative markers such as EST-SSR 029 and SSRCa 026 was crucial for accurately assessing genetic diversity and guiding breeding programs aimed at improving Robusta coffee's genetic base. The details about SSR marker characterization were summarized in Table 2.

Genetic diversity and DNA profiling of Robusta germplasm

The UPGMA dendrogram illustrated the genetic relationships among the Robusta germplasm accessions based on 11 polymorphic SSR markers (Figure 1). The clustering revealed three distinct groups of red, pink, and blue, indicating varying levels of genetic diversity within the germplasm, effectively distinguishing both closely related and genetically distant accessions. All accessions were successfully differentiated using these markers. The red group consisted of a single accession (62R), while the pink and blue groups contained 8 and 23 accessions, respectively. The pairwise genetic distance calculated using the 11 SSR markers ranged from 0.0 to 0.857 with the highest distance observed between accessions 35R and 122R. The development of an SSR fingerprinting panel aimed to provide an effective and cost-efficient approach to cultivar identification. To optimize the panel, SSR markers were selected based on their PIC values with five highly polymorphic markers (PIC > 0.6) chosen for their high discriminatory power. The genetic distance analysis using these markers

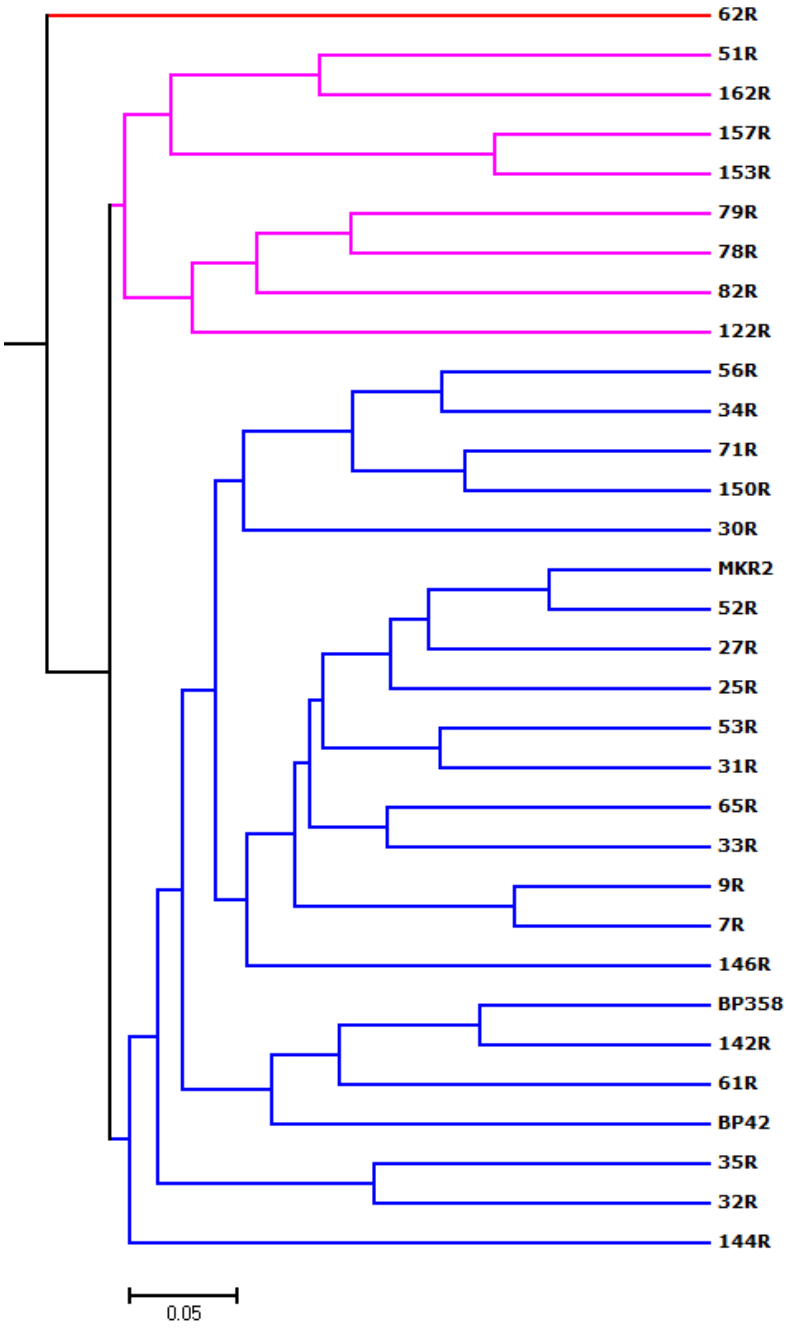

Figure 1. UPGMA dendrogram based on 11 polymorphic SSR markers.

demonstrated greater discrimination power compared to the analysis using 11 polymorphic SSR markers. Specifically, the pairwise genetic distance using 11 SSR markers ranged from 0.0 to 0.857, whereas the developed panel achieved a range from 0.0 to 1.0, enhancing the identification of accessions within the Robusta germplasm. This optimized panel not only improved accession differentiation but also reduced the overall cost and time required for analysis (Figure 2).

Previous studies have reported on DNA fingerprinting in Coffea using various DNA

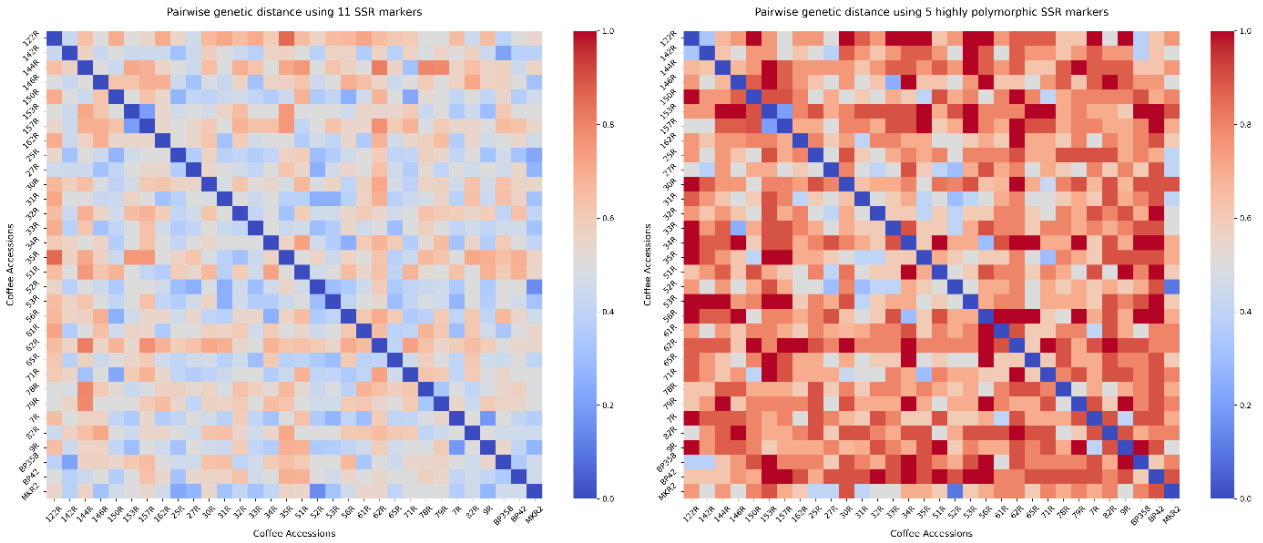

Figure 2. Heatmap of pairwise genetic distance of robusta accessions based on 11 SSR markers (left panel) and 5 highly polymorphic SSR markers (right panel).

markers [14, 15]. However, the plant materials or cultivars differed from those used in the present study. An SSR fingerprinting panel was developed to characterize Robusta accessions that were difficult to identify using morphological characteristics. This panel offered a superior approach to accession identification, enabling the process to be conducted even at the seedling stage, unlike morphology-based methods that required plants to mature. Additionally, the characterization process was not influenced by environmental factors or agronomic practices. Therefore, the SSR fingerprinting panel that included high-PIC-value SSR markers was the most effective method for characterizing accessions and evaluating genuine planting materials.

References

1. Barbosa WR, Souza FM, Lima LP. 2019. Genetic diversity of *Coffea canephora* using molecular markers: Implications for coffee improvement programs. Plant Mol Biol Rep. 37(4):249-258.
2. Diniz RP, Costa LS, Mendes PA. 2020. Genetic diversity in *Coffea canephora* revealed by microsatellite markers: A review of advances and future directions. Crop Sci. 60(2):334-345.
3. Faye A, Kane NA, Ndir KN. 2021. Genetic diversity and population structure of *Coffea canephora* in West Africa using microsatellite markers. Tree Genet Genomes. 17(1):89-98.
4. Nguyen QT, Tran TT, Vu HN. 2020. Assessment of genetic diversity of Robusta coffee germplasm in Vietnam using SSR markers. Plant Breed. 139(3):242-250.
5. Silva RM, Martins M, Pereira L. 2017. Genetic diversity in *Coffea canephora* assessed using microsatellite markers: Implications for conservation and breeding. Plant Breed. 136(2):221-229.
6. Mace ES, Buhariwalla KK, Buhariwalla HK, Crouch JH. 2003. A high-throughput DNA extraction protocol for tropical molecular breeding programs. Plant Mol Biol Rep. 21(4):459-460.
7. Missio RF, Caixeta ET, Zambolim EM, Zambolim L, Cruz CD, Sakiyama NS. 2010. Polymorphic information content of SSR markers for *Coffea* spp. Crop Breed Appl Biotechnol. 10(1):89-94.
8. Poncet V, Hamon P, Minier J, Carasco C, Hamon S, Noirot M. 2004. SSR cross-amplification and variation within coffee trees (*Coffea* spp.). Genome. 47(6):1071-1081.
9. Schuelke M. 2000 An economic method for the fluorescent labeling of PCR fragments. Nat Biotechnol. 18:233-234.
10. Arif IA, Khan HA, Shobrak M, Al Homaidan AA, Al Sadoon M, Al Farhan AH, *et al*. 2010. Interpretation of electrophoretograms of seven microsatellite loci to determine the genetic diversity of the Arabian Oryx. Genet Mol Res. 9:259-265.
11. Liu K, Muse SV. 2005. PowerMarker: An integrated analysis environment for genetic marker analysis. Bioinformatics. 21:2128-2129.
12. Kumar S, Stecher G, Tamura K. 2016. MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol Biol Evol. 33(7):1870-1874.
13. Botstein D, White RL, Skolnick M, Davis RW. 1980. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am J Hum Genet. 32(3):314–331.

14. Combes MC, Joët T, Lashermes P. 2018. Development of a rapid and efficient DNA-based method to detect and quantify adulterations in coffee (Arabica versus Robusta). Food Control. 88:198-206.
15. Pruvot-Woehl S, Krishnan S, Solano W, Schilling T, Toniutti L, Bertrand B, *et al*. 2020. Authentication of *Coffea arabica* varieties through DNA fingerprinting and its significance for the coffee sector. J AOAC Int. 103(2):325-334.